基于 HVS 的视频监控目标提取方法研究 *

赵 潇[1]，刘 畅[2,3,4]，田 霖[2,3]，韩 雪[2,3]，周继华[5]，吴 坚[5]

(1. 重庆邮电大学，通讯与信息工程学院，重庆 400065; 2. 移动计算与新型终端北京市重点实验室，北京 100190; 3. 中国科学院计算技术研究所无线通信技术研究中心，北京 100190; 4. 中国科学院大学，北京 100190; 5. 重庆金美通信有限责任公司，重庆 400000)

摘 要：目标识别是实现视频监控智能分析的基础，但在光照、阴影以及杂乱背景等场景中，往往会出现目标误判以及不合理聚类等问题。针对上述问题，提出一种基于人类视觉系统（HVS）的视频监控目标提取方法，结合 HVS 视觉关注原理，优化背景差法检测结果中存在的重复检测和错误分割问题，并根据 HVS 的跟踪特点以及目标运动的连续性，结合相邻帧检测结果，达到目标区域的完整准确提取；最后，基于实际采集视频进行仿真实验，证明所提目标检测算法结果准确性更高，在复杂背景下也有良好的检测效果。

关键词：监控系统；目标提取；背景差法；人类视觉系统

中图分类号：TP391.41 doi: 10.3969/j.issn.1001-3695.2018.01.0118

HVS-based object detection algorithm for surveillance video

Zhao Xiao[1], Liu Chang[2,3,4], Tian Lin[2,3], Han Xue[2,3], Zhou Jihua[5], Wu Jian[5]

(1. *School of Communications & Information engineering*, *Chongqing University of Posts & Telecommunications*, *Chongqing* 400065, *China*; 2. *Beijing Key Laboratory of Mobile Computing & Pervasive Device*, *Beijing* 100190, *China*; 3. *Wireless Center of Institute of Computing Technology*, *Chinese Academy of Sciences*, *Beijing* 100049, *China*; 4. *University of Chinese Academy of Sciences*, *Beijing* 100190, *China*; 5. *Chongqing Jinmei Communication Co*, *Ltd*, *Chongqing* 400000, *China*)

Abstract: Object detection is the basis of intelligent analysis, however, in the scene of illumination, shadow and clutter background, the problems of object misjudgment and unreasonable clustering is often appeared. Aiming at the above problems, propose a HVS-based object detection algorithm, which can optimize the error judgment and segmentation, and then accorded to the tracking characteristic of HVS and the continuity of object movement, combine the detection results of adjacent frames to achieve completely and accurately extraction object area. Finally, the simulation experiment based on the actual acquisition videos show that the proposed algorithm is more accurate and have good effect and in complex background.

Key words: surveillance system; object detection; background difference method; human visual system

0 引言

信息技术发展日新月异[1~3]，智慧地球、智慧城市的实现成为可能。随着智慧城市的快速发展，视频监控业务步入数据的井喷时代。根据 IDC（International Data Corporation 国际数据公司）的研究报告表明，2012 年视频监控数据总量约为 2.8 ZB，且正在以 55%的速度逐年增长，预计到 2020 年，数据总量约为 40 ZB，未来网络将难以承受。面对如此海量的数据，继续依靠传统的人工分析已经不能满足需求，结合智能算法分析的智能视频监控成为未来发展的必然趋势。

智能视频监控系统是指在系统中增加智能算法模块，对视频中运动目标区域进行定位、识别和跟踪，并在此基础上分析和判断目标的行为，识别监控系统中的异常情况，并以最快和最佳的方式发出警报或触发其他动作，从而有效进行事前预警、事中处理、事后及时取证的全自动、全天候、实时监控的智能系统。运动目标识别是实现智能分析的基础，决定了能否通过智能算法进行进一步的跟踪与监控[4,5]。

目前最常见的运动目标检测方法有三种：光流法、帧差法、背景差法。光流法不需要指定背景区域就可以实现对运动目标的检测，但计算量较大且对噪声较为敏感。帧差法对时间上连

收稿日期：2018-01-24；**修回日期：**2018-03-14 **基金项目：**北京市自然科学基金资助项目（L172049）；北京市青年拔尖人才资助项目（2015000021223ZK31）

作者简介：赵潇（1993-），女，河南南阳人，硕士研究生，主要研究方向为目标识别、目标提取（zhaoxiao@ict.ac.cn）；刘畅（1986-），男，博士，主要研究方向为无线通信、视频传输；田霖（1980-），女，副研究员，博士，主要研究方向为绿色无线通信系统、无线资源管理技术；韩雪（1981-），女，研究员，博士，主要研究方向为移动通信网络优化、网络流量管控；周继华（1979-），研究员，博士，主要研究方向为无线通信、移动互联网；吴坚（1985-），硕士，主要研究方向为无线网络、应急通信.

续的两帧或多帧图像进行差分运算，实现目标检测功能。背景差法首先利用监控视频流中的图像帧进行背景建模，然后将当前帧与背景帧进行差分运算，获取前景目标[6]。

视频监控具有背景相对稳定、感兴趣目标明确、帧间运动具有连续性以及帧间冗余较大的特点，同时考虑上述三种方法的原理、检测效果以及计算复杂度，认为背景差法最适合用于监控场景中进行目标识别。近年来，为提高目标检测精度而提出的优化算法有很多，通过调研分析，主要分为两个方向：1、优化目标检测算法；2、优化背景建模算法。其中优化目标检测算法主要有：文献[7~10]通过将背景差法分别与边缘检测、多帧差法等算法结果相结合，解决背景差法检测结果不完整、目标区域中存在空洞等问题；通过可变权重实现有效的变化检测，提高相对有效目标区域的权重，减少伪目标提取[11]。优化背景建模算法主要有：总结 HSV 颜色特征，文献[12]从图像序列中提取背景帧；文献[13]通过像素的概率密度对当前背景帧进行更新，解决检测不准确等问题；为了消除环境影响，文献[14]提出对图像分块，从而更新部分背景；文献[15]通过改进 Surendra 背景建模方法获取干净背景，达到抑制噪声和消除伪目标的目的；结合帧差法和曼哈顿距离，文献[16]提出优化背景初始化和目标检测算法。

然而，通过对上述文献研究发现，主要存在两个问题：上述方法大部分并没有考虑复杂背景的这一特殊场景，没有屏蔽在目标识别时复杂背景带来的影响；而且仅仅停留在目标识别阶段，没有考虑在目标分割提取时的问题。在动态变化场景中，如光照、阴影及杂乱背景的干扰，使感兴趣目标区域识别和聚类出现错误和不合理现象，如下图 1 所示。分析多帧检测结果，总结在目标提取时，背景差法检测结果存在问题如下：a) 提取出的感兴趣目标区域，存在大量由于噪声产生的伪目标；b)提取出的感兴趣目标区域，存在重复目标区域；c)提取出的感兴趣目标区域，存在错误分割、不合理聚类等问题。

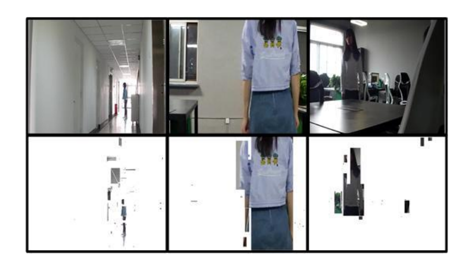


图 1 光照等引起的误检测

当感兴趣目标识别出现上述不合理结果时，对智能监控系统后期用到的智能算法,例如目标匹配、姿态识别、目标跟踪等的准确性带来较大影响，同时会带来巨大的计算冗余，影响系统分析结果。

研究表明，人类视觉总能快速定位出重要的目标区域并进行细致的分析，人类识别一副图像中的感兴趣区域，仅仅需要 13 ms，而对其他区域仅仅进行粗略分析甚至忽视。这种选择性的心理活动被称为人类视觉系统（human visual system, HVS）中的视觉关注机制，该机制通过对输入的外界信息进行有效的筛选，迅速将注意力集中到感兴趣区域[17]。这一机制体现了人类视觉系统主动选择关注内容并加以集中处理的视觉特性，该特性能有效提升图像内容筛选、目标识别等图像处理能力。因此，本文提出一种基于 HVS 的视频监控目标提取方法，具体研究内容包括：结合 HVS 视觉关注原理，首先分析背景差法检测结果，粗略定位出目标位置，判断是否存在细碎目标以及重复提取区域并进行优化；然后分析检测结果中是否存在不合理以及错误分割聚类，针对相应问题进行优化；然后根据 HVS 对目标的“跟踪”特性，设计了一种结合相邻帧检测结果的纠错方法；最后，对于视频流中的每一帧图像，实现上述优化过程，达到感兴趣目标区域的完整准确提取。

1 基于 HVS 的优化算法

在结合 HVS 视觉原理对感兴趣区域进行判断优化之前，首先通过背景差法得到初步检测识别的目标区域信息，包括目标信息的坐标以及大小等。

鉴于背景差法检测结果中存在的错误检测和分割的问题，首先，提出 HVS-intra-Frame（基于 HVS 的帧内）优化算法，该算法首先通过分析噪点特征，对检测结果中进行初步过滤删除，然后结合人类对场景中密集区域关注程度更高，通过 k-means 算法定位出目标区域的粗略位置，优化检测结果中存在的重复检测和错误分割的问题。最后，提出 HVS-inter-Frame（基于 HVS 的帧间）优化算法，该算法根据 HVS 的“跟踪”特点以及目标运动的连续性，结合相邻帧检测结果，进行进一步的纠错分析与优化。

1.1 HVS-intra-Frame 优化算法

在视频监控这一特殊场景中，感兴趣目标区域较为明确，结合视觉关注原理，可以有效识别出在复杂场景中由于光影造成检测结果中出现的细碎目标和噪声，从而实现对该类错误识别目标的过滤。HVS 视觉关注特点表明，在观察图片帧时，注意力会集中到场景中的密集区域，可以结合该原理检测并优化检测结果中的错误分割和不合理聚类问题。针对在复杂背景时，背景差法检测感兴趣目标区域出现的问题，本文设计优化算法流程如图 2 所示。

a)在复杂背景场景中，背景差法得到的感兴趣目标区域中包含噪点，会给后续智能算法的实现造成大量的计算冗余，需要进行初步的筛选和删除。分析大量噪点特性，具体表现为区域面积较小，因此，根据长期测试，以及分析目标区域和噪点的特征，设置合理门限值$thres1$，在得到背景差法识别出的目标区域位置信息之后，假设得到区域的宽和高分别为$width$和$height$，通过遍历初步得到的所有目标区域位置信息，如果满足

$$\begin{cases} width < thres1 \\ height < thres1 \end{cases} \tag{1}$$

则认为该区域为复杂背景中的噪声所引起的噪点，根据 HVS 视

觉关注原理以及监控视频中感兴趣目标明确的特点，判断该类目标不属于感兴趣目标区域，从而过滤删除目标区域中的噪点和细碎目标区域。

b)由于在测试过程中，发现在复杂背景时，感兴趣目标区域可能会被错误分割成若干区域，根据 HVS 视觉关注对密集区域关注程度更高，结合聚类算法，粗略定位出目标大致位置。首先求出得到的各个目标区域的中心点(x_c, y_c)，利用 k-means 算法对错误分割的若干区域中心点进行聚类，该算法首先随机选取若干目标中心点作为初始聚类中心$\mu_1, \mu_2, \ldots, \mu_k \in R^n$，主要分为以下两步:

(a) 对每一个聚类中心，计算剩余所有目标区域中心点到该聚类中心的距离，选出距离该聚类中心最近的几个样本作为一类：

$$c^{(i)} = argmin\|x^{(i)} - \mu_j\|2 \quad (2)$$

其中$x^{(i)}$为当前目标区域的中心点集。

(b)针对 a 中得到的类，根据每一类的目标位置，重新估计该类的中心：

$$\mu_j = \frac{\sum_{i=1}^{m} 1\{c^{(i)}=j\}x^{(j)}}{\sum_{i=1}^{m} 1\{c^{(i)}=j\}} \quad (3)$$

其中:m 代表初步得到的目标数目，1 代表真值判断，重复上述过程直至收敛。实现将目标区域划分为 L 簇，则 L 为当前图像中目标数目，完成图像中目标区域的位置和数量粗定位。如图 3 所示，图中蓝色区域代表根据背景差法提取到的目标区域，“红点”代表通过聚类得到目标区域大致位置。

获取目标信息
目标是否为噪点
是
删除该目标
否
聚类初步判断目标位置
目标是否存在包含
是
删除该目标
否
是否错误分割提取
否
结束
是
部分重叠
错误分割
实现合理提取

图 2 优化算法流程图



图 3 聚类算法结果

c)结合 HVS 视觉关注机制，判断感兴趣目标区域应该具有密集性，基于此特性，对初步识别结果中的不合理结果进行优化。分别遍历 k-means 算法得到的每一个簇中包含的目标位置信息，假设第i个目标坐标位置为x_i和y_i，宽和高分别为$width_i$和$height_i$，簇中第j个目标的位置信息分别为x_j和y_j，宽和高分别为$width_j$和$height_j$。本文在测试过程中发现，关于感兴趣目标区域的重叠，存在两种情况：1、两个目标区域完全重叠，即存在一个区域包含另一个区域的关系；2、两个目标区域部分重叠，即存在一个目标区域的部分属于另一个目标区域的关系。针对上述两种情况，可以根据位置信息进行判断，如果存在

$$\begin{cases} x_i \le x_j \\ y_i \le y_j \\ x_i + width_i \ge x_j + width_j \\ y_i + height_i \ge y_j + height_j \end{cases} \quad (4)$$

的关系，则说明目标j被i完全包含，即在识别结果中出现了目标重复检测的结果，应当对目标j进行删除。同时，如果感兴趣目标位置信息符合

$$|(2x_i + width_i - 2x_j - width_j)| \le |width_i + width_j| \quad (5)$$

$$|(2y_i + height_i - 2y_j - height_j)| \le |height_i + height_j| \quad (6)$$

则可以判断出检测结果中存在两个目标区域部分重叠的问题，通过对目标位置信息进行计算修正，重新聚类得到将两个区域包含在内的感兴趣区域。

d)判断错误分割目标位置信息，设置阈值$thres2$，遍历得到的感兴趣目标区域位置信息，满足

$$\begin{cases} x_i + width_i - x_j \le thres2 \\ y_i + height_i - y_j \le thres2 \end{cases} \quad (7)$$

时，说明目标区域被不合理识别聚类。分析两个感兴趣目标区域的位置信息，将不合理分割造成的结果合并聚类，达到完整提取感兴趣区域的目的。

1.2　HVS-inter-Frame 优化算法

研究表明，HVS 并非只简单理解进入人眼的视觉信号，而是存在一套内在的推导机制去解读输入的视觉信号，即对于待识别的场景，HVS 会根据大脑中的记忆信息，来推导、预测其视觉内容，实际上，人眼对运动物体的分辨能力和人眼能不能“跟踪”有关[18]。同时，基于监控视频场景相对固定的特点，监控视频具有连续性，则出现在场景中的运动目标也具有一定的连续性，可以对相邻帧检测结果进行综合分析，进一步提高准确性。结合视频监控中视频流的连续性，即感兴趣运动区域在相邻帧之间不会出现巨大的跳变这一原理，对上述感兴趣目标区域检测结果进行帧间纠错。

假设当前图像中目标数目为L_k，相邻上一帧中目标数目为L_{k-1}，如果$L_k = L_{k-1}$，则目标检测正确；如果$L_k > L_{k-1}$，包含两种情况，a.新目标出现，b.目标错误分割；如果$L_k < L_{k-1}$，包含两种情况:目标离开;目标丢失。

当$L_k > L_{k-1}$时，假设前一帧中检测到的目标区域为 Ω，遍历当前图像中检测得到的目标区域，如果当前图像中目标区域与 Ω 有交集，则计数为s个，当$s > 2$时，将当前图像中与上一帧

中有交集的目标区域合并，此时认为出现这种问题的原因是由于目标区域被错误的分割，并将结果与上一帧对应区域求并集，同时解决目标其他部分可能出现的缺失情况。

当$L_k < L_{k-1}$时，选取当前帧相邻的n帧联合分析，假设当前帧为第i帧，识别出的目标区域宽和高分别为$width$和$height$，本文通过判断相邻帧之间的目标区域面积之间的关系来判断当前帧检测到的感兴趣区域是否发生突变。具体过程如图 4 所示，根据长期测试以及目标变化的规律，设置阈值$thres3, thres4$，当

$$\begin{cases} \dfrac{width_i * height_i}{width_j * height_j} \geq thres3 \\ \dfrac{width_i * height_i}{width_j * height_j} \leq (1 - thres3) \end{cases} \quad (8)$$

时，说明当前帧i中对应的目标出现突变。记录出现突变的次数为l，当前帧i与相邻的n帧之间总对比的次数为m，在所有目标与相邻帧之间对应目标对比结束之后，得到总对比次数m和总出错次数l，然后通过计算l/m的值，如果$l/m \geq thres4$，则说明当前帧i中存在目标区域突变的情况。根据对感兴趣目标区域运动特性的分析，目标的运动具有连续性和均匀性，所以当当前帧出现目标突变之后，为了保证检测到的目标区域呈现出一致性和连续性，将当前帧的上一帧$i-1$帧得到的目标区域位置信息对当前帧进行赋值。

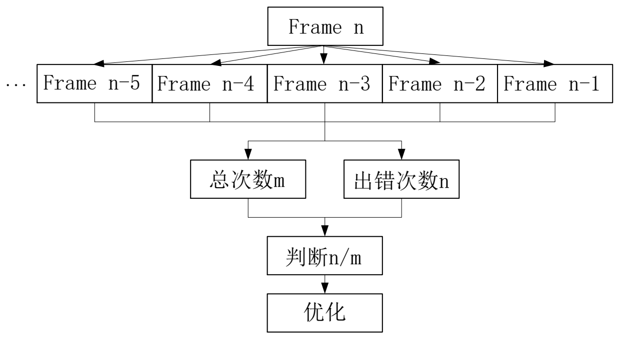


图 4 目标突变优化方法

2 实验

本文所选取实验环境以及实时采集视频参数分别如表 1 和 2 所示。

表 1 实验环境

CPU	内存	软件版本
Intel Core i7	8G	VS2013+OpenCV3.0

表 2 实时采集视频数据

场景	帧率
室内、走廊真实场景	24

本文的仿真是建立在背景差法的基础之上，通过长期实验测试，如图 5 所示，图中分别设置门限值为 5、15、25，可以看出，当$thres1$=5 时，仍然存在由于光照引起的噪点；当$thres1$=15 时，基本消除场景中的噪点；当$thres1$=25 时，出现目标区域大量丢失，因此设置合理门限值$thres1$为 15，将长或宽小于$thres1$的目标区域删除。如图 6 所示，当$thres2$=25 时，可以看出仍然存在目标区域被错误分割的情况，当$thres2 = 50$，目标区域可以被完整提取，因此设置门限值$thres2$为 50，将距离小于门限值$thres2$的目标区域合并为一个目标区域。如图 7 所示，通过对若干连续运动的目标区域面积进行分析，纠错算法中设置门限值$thres3$和$thres4$分别为 0.8 和 0.9。

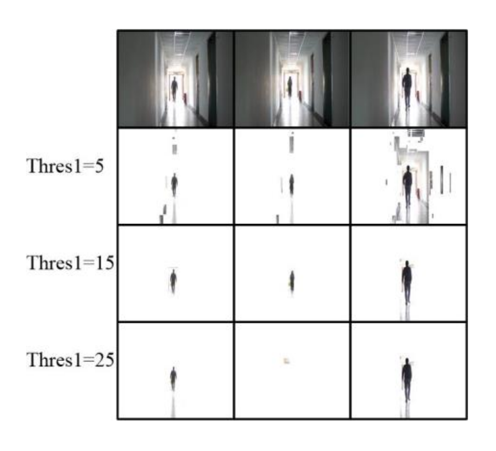


图 5 $thres1$选取

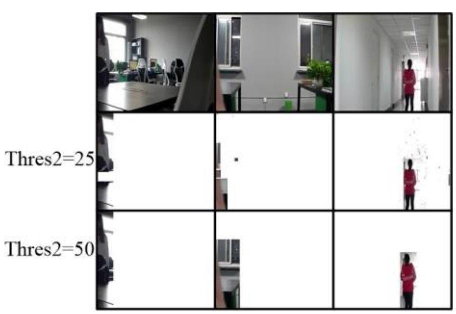


图 6 $thres2$选取

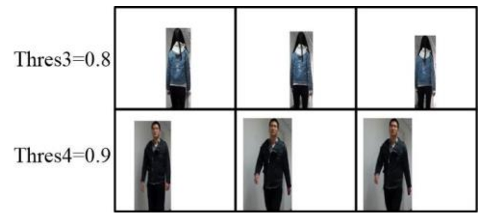


图 7 $thres3$和$thres4$选取

如图 8 所示，对实时采集的三种不同分辨率下各三段视频进行算法测试，所选场景为室内和走廊两种，每段视频时长在两小时以上，帧率为 24，处理图像帧数超过十万帧，图中横轴代表所选取视频图像分辨率，纵轴代表每秒处理图像帧数，仿真实验证明，在 640*480 分辨率下，平均每秒处理图像帧数为 34 帧，800*600 分辨率下，平均每秒处理图像帧数为 28 帧，1280*720 分辨率下，平均每秒处理图像帧数为 22 帧。因此，本算法在视频监控场景下可以满足实时视频处理。

在相同实验环境下，本文分别选取三种场景下各 1000 帧图片进行测试，如图 9 所示，横坐标表示所选取的存在动态变化的三种场景，包括光照、阴影和杂乱背景三种，纵坐标表示准确检测目标图像帧数所占总帧数的百分比，图中三种柱状图分别表示背景差法、Intra Frame 算法与 Intra Frame＋Inter Frame 算法的检测结果。从图中可以看出，本文检测算法相对于背景

差法检测结果准确率可以改进约 60%以上。

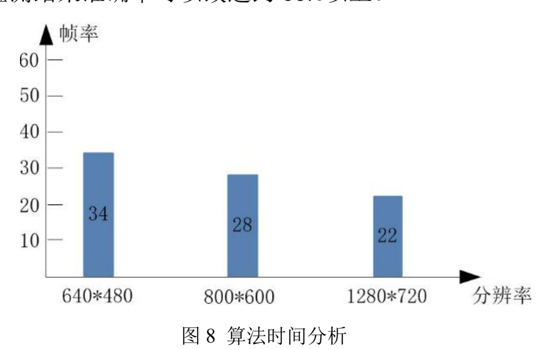


图 8 算法时间分析

另外，测试环境不变，选取三种存在动态场景变化的测试结果分别如下图 10、11、12 所示，场景中椭圆区域内为环境引起的噪点或目标区域出现错误分割的部分。根据仿真结果可以看出，在受到场景的动态变化，如光照、阴影及杂乱背景的干扰等的影响的时候，背景差法的检测结果会出现大量噪点以及不完整提取，本文优化算法相对于背景差法可以更加完整准确的检测出场景中的感兴趣目标区域，同时对场景中出现的噪点等可以有有效的抑制。

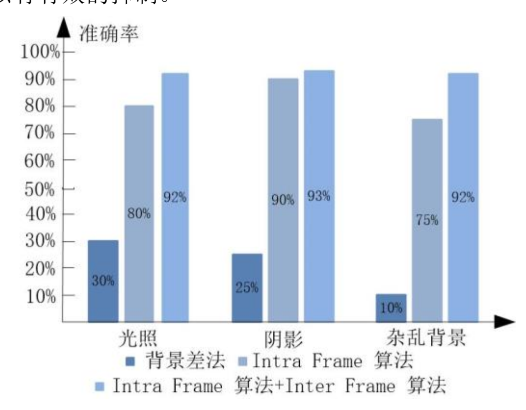


图 9 性能对比

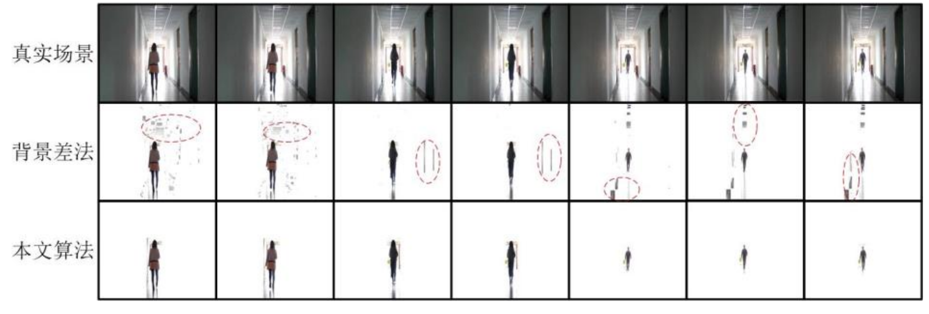


图 10 场景 1

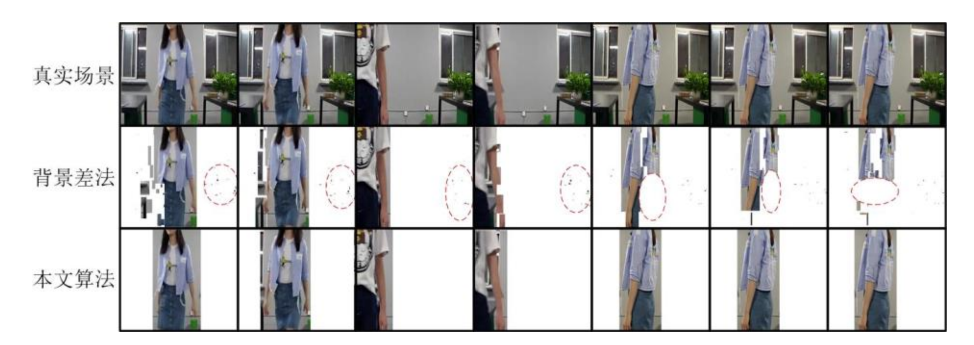


图 11 场景 2

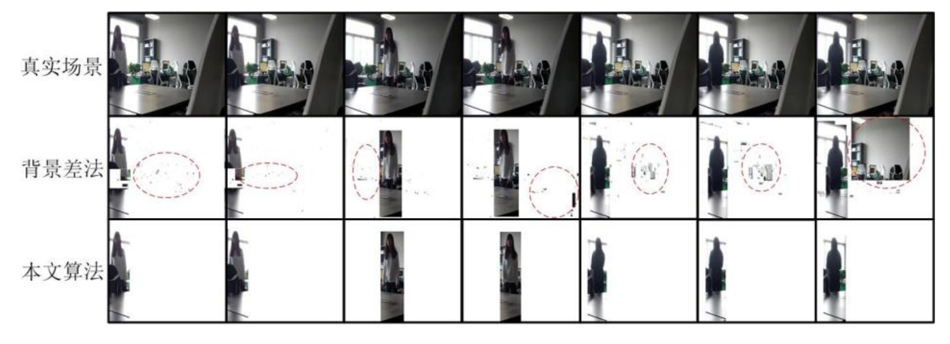


图 12 场景 2

3 结束语

本文提出了一种基于 HVS 的视频监控目标提取方法。该方法通过将 HVS 中的视觉关注特点引入背景差法，快速优化目标检测中的错误判断和分割，从而完整准确提取视频监控场景中的前景目标，有效弥补上述所提现有算法的不足。首先通过帧内聚类优化解决错误检测和分割的问题，进一步通过帧间检测结果分析判断优化帧间突变问题，达到有效抑制环境影响目标检测结果的目的。仿真结果表明，本文改进的算法相比于传统的背景差法，具有更好的检测效果，可以准确完整的实现目标区域提取。

参考文献：

[1] Liu Ling, Zhou Yiqing, Garcia V, et al, Load aware joint CoMP clustering and inter-cell resource scheduling in heterogeneous ultra dense cellular networks [J]. IEEE Trans on Vehicular Technology, 2017, 99 (11): 1-14.

[2] Zhou Yiqing, Liu Hang, Pan Zhengang, et al, Two-stage cooperative multicast transmission with optimized power consumption and guaranteed coverage [J]. IEEE JSAC on SEED, 2014, 32 (2): 274-284.

[3] Garcia V, Zhou Yiqing, Shi Jinglin, Coordinated multipoint transmission in dense cellular networks with user-centric adaptive clustering [J]. IEEE Trans on Wireless Comm. , 2014, 13 (8): 4297-4308.

[4] Javed O, Rasheed Z, Alatas O, al. KNIGHT™: a real time surveillance system for multiple and non-overlapping cameras [C]// Proc of IEEE International Conference on Multimedia and Expo. 2003: 649-652.

[5] 梁何. 视觉显著性在视频目标识别中的应用研究 [D]. 南京：南京邮电大学, 2016. (Liang He. Research on applications of visual saliency in video object recognition [D]. Nanjing: Nanjing University of Posts and Telecommunications, 2016)

[6] Bang J, Kim D, Eom H. Motion object and regional detection method using block-based background difference video frames [C]// Proc of IEEE International Conference on Embedded and Real-Time Computing Systems and Applications. 2012: 350-357.

[7] Wang Tongyao. The motion detection based on background difference method and active contour model [C]// Proc of IEEE Joint International Information Technology and Artificial Intelligence Conference. 2016: 480-483.

[8] Li Xiaoping, Ren Xulong, Song Ye *et al.* Improved difference method for fast-moving object recognition in complex background [C]// Proc of International Conference on Information Technology and Electronic. 2014: 137-140.

[9] Lan Yongtian, Ji Zhijian, Gao Junwei, *et al.* Robot fish detection based on a combination method of three-frame-difference and background subtraction [C]// Proc of Chinese Control and Decision Conference. 2014: 3905-3909.

[10] Srivastav N, Agrwal S L, Gupta S K, *et al.* Hybrid object detection using improved three frame differencing and background subtraction, [C]// Proc of International Conference on Cloud Computing, Data Science & Engineering. 2017: 613-617.

[11] Jiang Shengqin, Lu Xiaobo. WeSamBE: a weight-sample-based method for background subtraction [J]. IEEE Trans on Circuits and Systems for Video Technology, 2017, PP (99): 1.

[12] Zhao Yabin, Lv Guoyun, Ma Tiantian, *et al.* A novel method of surveillance video Summarization based on clustering and background subtraction [C]// Proc of International Congress on Image and Signal Processing. 2015: 131-136.

[13] Yang Caiyun, Gao Jingmin, Chen Fubin, Embedded moving image detection based on background subtraction and finite difference method [C]// Proc of International Congress on Image and Signal Processing. 2012: 68-71.

[14] Bang J, Kim D, Eom H, Motion object and regional detection method using block-based background difference video frames [C]// Proc of IEEE International Conference on Embedded and Real-Time Computing Systems and Applications. 2012: 350-357.

[15] 王凯，吴敏，姚辉，等. 多帧背景差与双门限结合的运动目标检测方法 [J]. 小型微型计算机系统, 2017, 38 (1): 179-183. (Wang Kai, Wu Min, Yao Hui, *et al.* Moving target detection method based on multi frame background subtraction and double threshold [J]. Journal of Chinese Computer Systems, 2017, 38 (1): 179-183)

[16] 朱克佳，李俊. 曼哈顿距离的背景差法运动目标检测 [J]. 激光杂志, 2017, 38 (10): 85-89. (Zhu Kejia, Li Jun. Moving object detection of background subtraction method based on Manhattan distance [J]. Laser Journal, 2017, 38 (10): 85-89

[17] Ma Jianshe, Guo Libo, Su Ping. Visual saliency detection based on disperse degree of color [C]// Proc of International Conference on Multimedia and Image Processing. 2017: 38-42.

[18] Dong Yuanyuan, Pourazad M T, Nasiopoulos P. Human visual system-based saliency detection for high dynamic range content [J]. IEEE Trans on Multimedia, 2016, 18 (4): 549-562.